

# Gene Pool Subdivision of East African Sweetpotato Parental Material

Maria C. David, Federico C. Diaz, Robert O. M. Mwanga, Silver Tumwegamire, Roberto C. Mansilla, and Wolfgang J. Grüneberg\*

## ABSTRACT

Sweetpotato [*Ipomoea batatas* (L.) Lam] breeding is important for food security and health in East Africa (EA), and a breeding platform in Uganda provides national researchers and breeders in EA with true seed. Our objectives were to characterize genetic relationships among parental material used at the EA breeding platform. There were 135 parents and six check clones analyzed using 31 simple sequence repeat primers. An average of 7.13 alleles per primer was found, and Jaccard similarity coefficients were in the range of 0.298 to 1.00 with a mean of 0.542. Unweighted pair group cluster analysis placed most African parents in two main subclusters showing no association with morphology or geographical origin. The subclusters were also supported by principal coordinate analysis, derivative analysis of principal components, and population structure simulations. The analyzed breeding material from EA was highly genetically variable, grouped in two distinct genetic pools, and suitable to study heterosis exploiting breeding schemes.

M.C. David, F.C. Diaz, and W.J. Grüneberg, International Potato Center (CIP), P.B. 1556 Lima 12, Peru; R.O.M. Mwanga, CIP, P.B. 22274, Kampala, Uganda; S. Tumwegamire, International Institute of Tropical Agriculture (IITA), P.B. 34441, Dar-es-Salaam, Tanzania; R.C. Mansilla, Univ. Nacional Agraria La Molina, P.B. 12-056, Lima 12, Peru. Received 26 Nov. 2017. Accepted 26 Feb. 2018. \*Corresponding author (w.gruneberg@cgiar.org). Assigned to Associate Editor Dina St. Clair.

**Abbreviations:** AMOVA, analysis of molecular variance; CC, check clone; DAPC, discriminant analysis of principal components;  $D_L$ , discriminating power; EA, East Africa; FV, farmer varieties; HEBS, heterosis exploiting breeding scheme(s); MV, modern varieties; OFSP, orange-fleshed sweetpotato; PC, principal component; PCoA, principal coordinates analysis; PCR, polymerase chain reaction; PIC, polymorphic information content; SPVD, sweetpotato virus disease; SSA, sub-Saharan Africa; SSR, simple sequence repeat; UPGMA, unweighted pair group method analysis.

**SWEETPOTATO** [*Ipomoea batatas* (L.) Lam] is a major food crop in sub-Saharan Africa (SSA). It is propagated by cloning through vine cuttings. In East Africa (EA), annual production reached 10.4 Tg in 2014 (FAOSTAT, 2018; accessed for Burundi, Ethiopia, Kenya, Rwanda, Tanzania, and Uganda). The crop is grown for food security and nutritional goals. White- and cream-fleshed storage roots are usually consumed; however, for the last two decades there has been interest in introducing orange-fleshed sweetpotato (OFSP) to alleviate vitamin A deficiency (Low et al., 2001). Additionally, sweetpotato leaves are used as a vegetable, and aboveground biomass has become important as animal feed in the past decade (Low et al., 2009). Efficient sweetpotato breeding programs are important to increase food security and improve health in SSA, and a basic aspect of breeding is the characterization of crossing parents.

Varietal development is straightforward in sweetpotato because each true seed plant already represents a potential variety. Selection

Published in Crop Sci. 58:1–13 (2018).  
doi: 10.2135/cropsci2017.11.0695

© Crop Science Society of America | 5585 Guilford Rd., Madison, WI 53711 USA  
This is an open access article distributed under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

can be accelerated by reducing test years and increasing the number of locations. Population improvement is challenging because the crop is a highly heterozygous hexaploid hybrid. This nurtures the expectation that heterosis is important for sweetpotato performance, and that systematic exploitation of heterosis can improve the efficiency of population improvement (Grüneberg et al., 2015). True seed production occurs easily by open pollination, and a successful pollination results in one to four true seeds, so polycrosses have been the backbone of population improvement (Martin and Jones, 1986). Breeding in EA started in Rwanda and Uganda in 1983 and 1987, respectively, on the basis of polycrosses with 15 to 30 parents. In 2003, CIP and HarvestPlus (Pfeiffer and McClafferty, 2007) started an initiative to breed OFSP in SSA. Within this, National Agriculture Research Systems in Tanzania, Kenya, and Ethiopia started to breed sweetpotato. In 2008, the program at Namulonge in Uganda considerably increased the number of parents (increased to 140–150 parents) and became a CIP breeding platform for EA serving National Agriculture Research Systems with true seed supply, technical backstopping, and support in proposal writing. Since 2009, this platform has been supported by the Bill & Melinda Gates Foundation. The major traits of interest in EA are: (i) resistance to sweetpotato virus disease (SPVD, interaction of *Sweetpotato chlorotic stunt virus* and *Sweetpotato feathery mottle virus*; farmers usually propagate sweetpotato by cloning, without obtaining new seeds regularly), (ii) storage root yields and the number of commercial storage roots per plant, (iii) elevated pro-vitamin A with high root dry matter and starchy taste, (iv) abundant upper biomass production to facilitate vine production for seed systems and for use as animal feed, and (v) earliness and suitability for piece-meal harvest. For further details on breeding targets and an outline of seed systems in EA in linkage with new variety releases, refer to Mwanga et al. (2017).

Theoretically, population improvement by polycross breeding nurseries should be inferior to controlled cross breeding. However, polycrosses of sweetpotato can easily generate large amounts of true seed. In contrast, controlled crosses allow selecting of parents on the basis of offspring performance and to test and perhaps implement a heterosis exploiting breeding scheme (HEBS). The prerequisite for a HEBS is searching for or developing mutually heterotic gene pools in parental material (Melchinger, 1999). Moreover, a HEBS also allows enhanced inbreeding within gene pools, and this is desirable for sweetpotato breeding in EA because of resistance to SPVD, which is thought to be recessively inherited and occurs at very low frequencies (Mwanga et al., 2002a, 2002b). The pressure of SPVD is extremely high in EA. An intensive discussion has commenced concerning the use of HEBS in clonal breeding of the CGIAR programs (Miles, 2007; Grüneberg et al., 2009). Use of HEBS in clonal breeding was

proposed long ago (Hull, 1945; Melchinger and Gumber, 1998), but recommendations were applied to heterosis in traditional hybrid crops and not to population improvement of the Root, Tuber and Banana program in the CGIAR. Sugarbeet (*Beta vulgaris* L.) is an exception concerning use of HEBS in root crops (Bosemark, 2006), but this crop is propagated by true seed. There are attempts to use HEBS in potato (*Solanum tuberosum* L.) breeding (Lindhout et al., 2011; Jansky et al., 2016), but these are in experimental stages and have not been adopted by breeders. We hypothesize that the large number of sweetpotato parents in the crossing block at Namulonge in Uganda can be grouped into gene pools to serve as a basis to test and implement HEBS for sweetpotato breeding in EA.

The establishment of heterotic groups can be based on geographic origin, agronomic traits, pedigree data, or molecular data (Melchinger, 1999). Mutually heterotic gene pools can be developed in sweetpotato, as has been shown for two South American OFSP breeding populations, and simple sequence repeat (SSR) markers were very useful in this context (Federico Diaz, personal communication, 2014; results partially described in Grüneberg et al., 2015). Molecular markers have been demonstrated to be a useful tool to group parental material into gene pools for HEBS in many crops: maize (*Zea mays* L.; Messmer et al., 1992; Reif et al., 2005), rye (*Secale cereale* L.; Fischer et al., 2010), winter triticale ( $\times$  *Triticosecale* Wittm.; Tams et al., 2004), oilseed rape (*Brassica napus* L.; Becker et al., 1995), faba bean (*Vicia faba* L.; Link et al., 1995), rice (*Oryza sativa* L.; Xiao et al., 1996) and sorghum [*Sorghum bicolor* (L.) Moench; Bhosale et al., 2011]. African sweetpotato germplasm has been repeatedly characterized as having high morphological diversity in EA (Kaledzi et al., 2010; Yada et al., 2010a; Elameen et al., 2011). Agronomic traits indicated high diversity in local EA farmer varieties (FVs) (Tumwegamire et al., 2011a).

With respect to molecular markers, sweetpotato germplasm from SSA has been investigated using amplified fragment length polymorphism (Elameen et al., 2008) and SSR markers (Gichuru et al., 2006; Yada et al., 2010b; Tumwegamire et al., 2011b). These studies predominantly used FVs and not the entire parental material of a breeding program. Analysis of FVs from EA ( $N = 266$ ) with four microsatellite markers showed no relationships between geographic origin and clustering of clones (Gichuru et al., 2006). For 94 Tanzanian cultivars analyzed using amplified fragment length polymorphism markers, there were no correlations between morphological and molecular data (Elameen et al., 2008). High genetic diversity was detected for 192 Ugandan germplasm accessions characterized by 10 microsatellite markers, but there was no relationship between the district of clone origin and marker clustering (Yada et al., 2010b). In a study using 26 microsatellite markers to compare white-fleshed and

rarely found OFSP FVs from SSA, the OFSP from Africa was clearly distinct from OFSP from the Americas and closely related to white-fleshed sweetpotato from Africa (Tumwegamire et al., 2011b). Parental material used in EA breeding nurseries has not been studied using molecular markers for gene pool subdivision.

This study had two objectives. The first was to characterize genetic relationships among sweetpotato parents with microsatellite markers used by CIP's breeding platform in EA at the National Crops Resources Research Institute at Namulonge in Uganda. The second objective was to allocate parents of the Namulonge polycross crossing block into two potentially heterotic gene pools

and to identify potential cross combinations for further studies on HEBS.

## MATERIALS AND METHODS

### Plant Material

A total of 141 sweetpotato accessions were used for this study (Table 1) and described by clone name, local code, cultivar type, country of origin, plant type, root flesh and skin color, storage root form, and CIP and Ugandan germplasm codes (Supplemental Table S1). Of these, 135 accessions were parents from the polycross crossing block at Namulonge in Uganda, and six clones from the CIP genebank in Peru were used as checks. The parents were mainly of EA origin: 119 from EA,

**Table 1. Description of clones used for the genetic diversity study.**

Origin	No. of clones	Description by clone name, code, country of origin, variety type, and storage root flesh color
Parental material of East African origin	119	<p><u>Clone name and code:</u> K-118 (KE09), Oguroiwe (KE11), K-566632 (KE14), SPK004 (KE19), Ubuogo (KE21), Mugande (RW01), Carrot Dar (TZ01), Mayai (TZ02), Carrot C (TZ03), Ukerewe (TZ04), Magabari (UG05), NN (UG06), Karebe (UG15), Kigabali (UG19), Kyebandula (UG20), Tororo 3 (UG23), Osapat (UG29), Kala (UG40), Abuket 1 (UG41), Ejumula (UG43), Kamamanzi (UG44), NN (UG45), Wagabolige (UG47), Osukut (UG51), Opaade (UG52), Kakoba (UG53), Epura Amojong (UG54), Anyumel (UG55), Oleke (UG56), Kalobo (UG57), Rwabuganda (UG58), Kibogo (UG59), NK318L (UG60), NN (UG61), Kyebandira 2 (UG62), Anamoyito (UG63), Mary (UG64), Silk Omuyaka (UG65), Koromojo (UG66), Tedolo Kerani (UG67), Liralira (UG68), Kahungezi (UG69), Burundi (UG70), NN (UG71), Kalebe (UG72), Kibanda (UG73), NN (UG74), Bunduguza Empyaka (UG75), Dimbuka Obuleku (UG76), Dimbuka (UG77), Silk (UG78), Otada (UG79), NASPOT 1 (UG80), NASPOT 5 (UG81), Kyabafuluki (UG82), NASPOT 5/58 (UG83), Kampala Red (UG84), Silk (UG85), NASPOT 3 (UG86), NN (UG87), Dar-es-Salaam Carrot (UG88), Suwedi (UG89), NN (UG90), Tuulansime (UG91), Bungoma (UG92), Duduma 2 (UG93), NN (UG94), Dduka Enzala (UG95), Koromojo Red (UG96), Bikiramaia (UG97), NN (UG98), NN (UG99), NN (UG100), Oketodede (UG101), Nylon (UG102), NN (UG103), Uganda Mali (UG104), NN (UG105), Munafu Dimbuka (UG106), NN (UG107), Mugiga (UG108), Bunduguza 2 (UG109), Kigaire (UG110), Namusoga (UG111), NN (UG112), Koromojo Red (UG113), Bunduguz Empyaka 2 (UG114), Gulu (UG115), Dimbuka (UG116), New Kawogo (UG117), Bitambi (UG118), Sowola 389A (UG119), NN (UG120), NASPOT 7 (UG121), NASPOT 10 O (UG123), NASPOT 11 (UG124), NK259L (UG125), Semanda (UG126), NN (UG127), Mukoma (UG128), Silimu (UG129), NN (UG130), NN (UG131), NN (UG132), Dimbuka-Bukulula (UG133), Dagadaga (UG134), Kawogo Old (UG135), NN (UG136), Woluganda (UG137), BND145L (UG138), Tengerere (UG139), Mpaifumbiro (UG140), Mpambire (UG141), Tanzania (UG142), Silk (UG143), Malagalia (UG144), Mbale (UG145), and Tanzania (UG146).</p> <p><u>Country of origin:</u> Kenya, Rwanda, Tanzania and Uganda with a frequency of 5, 1, 5 and 108, respectively.</p> <p><u>Variety type:</u> breeding line, FV, and MV† with a frequency of 4, 105, and 10, respectively.</p> <p><u>Storage root fresh color:</u> white, cream, pale yellow, yellow, pale orange, orange, and dark orange with a frequency of 14, 68, 1, 18, 4, 9, and 4, respectively.</p>
Parental material of Southern African origin	1	<p><u>Clone name and code:</u> Rainha (MZ01).</p> <p><u>Country of origin:</u> Mozambique.</p> <p><u>Variety type:</u> FV.</p> <p><u>Storage root fresh color:</u> pale yellow.</p>
Parental material of West African origin	2	<p><u>Clone name and code:</u> TIS-9265 (CIP440076) and TIS-9101 (CIP440099).</p> <p><u>Country of origin:</u> Nigeria.</p> <p><u>Variety type:</u> MV.</p> <p><u>Storage root fresh color:</u> cream and orange.</p>
Parental material of non-African origin	13	<p><u>Clone name and code:</u> Santo Amaro (CIP400011), Dagga (CIP199062.1), Huarmeyano (CIP420020), Zapallo (CIP420027), Tainung 64 (CIP440189), Beauregard (CIP440132), Caromex (CIP440136), DLP3163 (CIP420269), Excel (CIP440016), Jewel (CIP440031), Resisto (UG), W-115 (CIP440424), and WT-237 (n.a.‡).</p> <p><u>Country of origin:</u> Brazil, Peru, Taiwan, USA, and n.a. with a frequency of 1, 4, 1, 6, and 1, respectively.</p> <p><u>Variety type:</u> FV and MV with a frequency of 2 and 11, respectively.</p> <p><u>Storage root fresh color:</u> cream, yellow, orange, and dark orange with a frequency of 1, 2, 8, and 2, respectively.</p>
Check clones	6	<p><u>Clone name and code:</u> Xushu 18 (CIP440025), Yanshu 1 (CIP440024), SPK004 (CIP441768), Naveto (CIP440131), Jonathan (CIP420014), and Resisto (CIP440001).</p> <p><u>Country of origin:</u> China, Kenya, Papua New Guinea, Peru, and USA with a frequency of 2, 1, 1, 1, and 1, respectively.</p> <p><u>Variety type:</u> FV and MV with a frequency of 2 and 4, respectively.</p> <p><u>Storage root fresh color:</u> white, cream, pale yellow, orange, and dark orange with a frequency of 1, 2, 1, 1, and 1, respectively.</p>

† FV, farmer variety; MV, modern variety.

‡ n.a., not available.

one from South Africa, two from West Africa, and 13 of non-African origin. Parents mainly comprised FVs (108 clones) and some modern varieties (MVs) or breeding lines (27 clones); 28 parents were OFSP, and 107 parents were non-OFSP. Check clones (CCs) were not part of the polycross recombination and included ‘Jonathan’ (FV from Peru), ‘Xushu 18’ and ‘Yanshu 1’ (MVs from China), ‘Naveto’ (FV from Papua New Guinea), ‘Resisto’ (MV from the United States), and ‘SPK004’ (FV from Kenya held in trust at CIP’s genebank in Peru). All CCs were obtained as plantlets from CIP’s genebank, whereas all parental material from the crossing block was obtained as lyophilized leaves from Uganda.

SSR Amplification

Lyophilized leaves were used to isolate total DNA by the method of Murray and Thompson (1980). DNA samples were quantified, and 60 ng of total genomic DNA from each sample was used per polymerase chain reaction (PCR). All reactions had a total volume of 10 µL containing 2.5 mM MgCl2, 0.2 mM deoxynucleotides, 22 nM forward primer, 15 nM reverse primer, 25 nM M13 Forward 700/800 (LI-COR), and 0.025 U µL<sup>-1</sup> Taq DNA polymerase (New England Biolabs), each as a final concentration. The thermocycler amplification program had a first step at 94°C that lasted 4 min, then 29 cycles as follows: denaturing at 94°C for 1 min, alignment for 1 min,

and extension at 72°C for 1 min. There was a final of extension step of 72°C for 7 min, and the temperature was then dropped to 4°C. Annealing temperatures for PCR varied according to the primer (Table 2). The PCR products were run in a polyacrylamide gel in a DNA Analyzer 4000 (LI-COR).

SSR Data Scoring and Analysis

Genotypes were scored for the presence (1) or absence (0) of bands visualized using the software Saga<sup>GT</sup> version 3.3. (LI-COR, 2005). The number of alleles per locus, percentage of polymorphic loci, and polymorphic information content (PIC) were calculated. The PIC was calculated with the equation

PIC = 1 - \sum\_{i=1}^n p\_i^2

where p<sub>i</sub> is the frequency of the i<sup>th</sup> allele (Weir, 1996). The limit of the discriminating power (D<sub>L</sub>) was calculated with the equation

D\_L = 1 - \sum\_{i=1}^n p\_i^2

where p<sub>i</sub> is the frequency of the i<sup>th</sup> haplotype (Kloosterman et al., 1993).

A similarity matrix was obtained by the Jaccard similarity coefficient and a dendrogram with the unweighted pair group

Table 2. Description of simple sequence repeat markers used to characterize sweetpotato genotypes used at the breeding platform in Uganda by currently used names, motifs, forward and reverse primers, and annealing temperature.

Name	Forward primers	Reverse primers	Motif	Temp.	Reference
	5'–3'			°C	
Ib-242	GCGGAACGGACGAGAAAA	ATGGCAGAGTGAAAATGGAACA	(CT) <sub>3</sub> CA(CT) <sub>11</sub>	58.0	Buteler et al. (1999)
Ib-286	AGCCACTCCAACAGCACATA	GGTTTCCCAATCAGCAATTC	(CT) <sub>12</sub>	57.0	Buteler et al. (1999)
Ib-297	GCAATTTCCACACAAACACG	CCCTTCTTCCACCATTTC	(CT) <sub>13</sub>	58.0	Buteler et al. (1999)
IBCIPI-1	CCCACCCCTTCATTCCATTACT	GAACAACAACAAAAGGTAGAGCAG	(ACC) <sub>7</sub>	63.0	Yañez (2002)
IbC5	CCACAAAAATCCAGTCAACA	AGTGGTCGTCGACGTAGGTT	(AAG) <sub>8</sub>	62.0	Solis et al. (unpublished data, 2008)
IbC12	TCTGAGCTTCTCAAACATGAAA	TGAGAATTCCTGGCAACCAT	(TTC) <sub>6</sub>	56.0	Solis et al. (unpublished data, 2008)
IbE2	CAGCCGCCAAGTTTCTACA	AGGCGGAGGCTGATAATGA	(TCT) <sub>13</sub>	62.0	Solis et al. (unpublished data, 2008)
IbJ67	CACCCATTTGATCATCTCAACC	GGCTCTGAGCTTCCATTGTTAG	(GAA) <sub>5</sub>	58.0	Solis et al. (unpublished data, 2008)
IbJ116A	TCTTTTGCATCAAAGAAATCCA	CCTCAGCTTCTGGGAAACAG	(GAA) <sub>8</sub>	57.0	Solis et al. (unpublished data, 2008)
IbJ263	CTCTGCTTCTCCTGCTGCTT	GTGCGGCACCTTGCTTTGATA	(AAC) <sub>5</sub>	55.5	Solis et al. (unpublished data, 2008)
IbJ522a	ACCCGCATAGACACTCACCT	TGACCGAAGTGATCTAGTGG	(CAC) <sub>6–7</sub>	57.0	Solis et al. (unpublished data, 2008)
IbJ544b	AGCAGTTGAGGAAAGCAAGG	CAGGATTTACAGCCCCAGAA	(TCT) <sub>5</sub>	62.0	Solis et al. (unpublished data, 2008)
IbJ664E	CACATGCCATGGACGCTCCAA	GATTCTTCTCCTTCCAGCTCCT	(CTT) <sub>6</sub>	55.0	Solis et al. (unpublished data, 2008)
IbN21	AACCCTCATCTTTCTCATCTCTTC	ACCTTGAACCTCCGTCTCCTCTT	(CT) <sub>0</sub> C	60.0	Huamani et al. (unpublished data, 2010)
IbN24	TAATGAGGTGTGATGATGGGTACTA	AGTGAAAGTTGAGTCAAGGAAAATC	(TA) <sub>5</sub> GA(TA) <sub>3</sub>	60.0	Huamani et al. (unpublished data, 2010)
IbN37	GATGATGGAGCTCATAAATCTCG	GTCACTGTGTCTCCAGTTTTTC	(TA) <sub>7</sub> T	55.0	Huamani et al. (unpublished data, 2010)
IBS144	TCGAACGCTTTCTACACTCTT	CTGTGTTTATAGTCTCTGGCGA	(TTC) <sub>9</sub>	60.0	Schafleitner et al. (2010)
IBS147	TGTGTACATGAGTTTGGTTGTG	GAAGTGCAACTAGGAAACATGA	(GCA) <sub>8</sub>	55.0	Schafleitner et al. (2010)
IBS149	CCACCTCCTTAGGTATCAGACT	ACTACTAGCGCTGCAACCTTAT	(AGA) <sub>8</sub>	60.0	Schafleitner et al. (2010)
IBS169	CGTACTATGTTTCCCCCATTAC	AATGCATCTACCCTCCTTACAC	(TTG) <sub>8</sub>	53.0	Schafleitner et al. (2010)
IBS199	TAAGTGGTTGCAGTGGTTTGT	ATAGGTCCATATACAATGCCAG	(ACA) <sub>7</sub>	60.0	Schafleitner et al. (2010)
IbY40	AGTGTGGGACTCATAAGATTCTG	GAATGAAATACAGTGACCCGAGAG	(GCG) <sub>7</sub> GC	60.0	Huamani et al. (unpublished data, 2010)
IbY44	CAAGAAGAGCATAAGCGTGAGAT	GCGATCTGAGAAGGTGATAATTG	(AGA) <sub>6</sub>	52.0	Huamani et al. (unpublished data, 2010)
IbY46	TAGTAACACCATTACTTATTAACCTTG	TGTAATCTCATGGATTGCTCGTAG	(ATC) <sub>5</sub> AT	55.0	Huamani et al. (unpublished data, 2010)
IbY51	GATGTCGTTTAGCGGACTGAG	GTATCGTCACATTACAGCAGCAG	(GCG) <sub>5</sub> G	55.0	Huamani et al. (unpublished data, 2010)
IbY52	AAACAGATAGCAGAGACGAGATGAG	CAGATAGTGTACCAACACTGAAGA	(GCG) <sub>5</sub> G	55.0	Huamani et al. (unpublished data, 2010)
IbY53	CCACGATCTCGGAAACCGCCAT	GGGGCAAAAGGCTCTTATTATAT	(GGA) <sub>5</sub> G	55.0	Huamani et al. (unpublished data, 2010)
IbY54	GTCCAAGAGAAAAGAACTGAAGATG	AACTATTCTGCACACTACATGCTC	(TGT) <sub>5</sub> T	57.0	Huamani et al. (unpublished data, 2010)
IbY56	CACCATGGATTTCAAACCACTACTT	AGGGGGAGTTGTCTTGACTGGT	(CCT) <sub>5</sub>	52.0	Huamani et al. (unpublished data, 2010)
IbY58	ACGACATGGCTCTCTCTTTCTC	AGTTTCCTTTCTCGACGCTTCT	(GCG) <sub>5</sub>	55.0	Huamani et al. (unpublished data, 2010)
IbY60	TCTCTCTGTTATGTTATGGTGATGG	GCGTTTTACAAGATTCAGAAACCAC	(TAT) <sub>5</sub>	62.0	Huamani et al. (unpublished data, 2010)



method analysis (UPGMA) using the software NTSYSpc version 2.2 (Rohlf, 2000). On the basis of the UPGMA results, a subdataset was created comprising parents of African origin with potential to be grouped into two gene pools (i.e., presence and absence data of 119 genotypes). On the subdataset, a principal coordinates analysis (PCoA) was performed using DARwin version 6 (Perrier and Jacquemoud-Collet, 2006). Next, the subdataset was recoded as alleles as requested by the “adeget” package (Jombart, 2008) and artificial alleles were eliminated using the “poppr” package (Kamvar et al., 2014) to allow a discriminant analysis of principal components (DAPC) using “adeget” in the software R (R Core Team, 2014). For the DAPC, four clusters were chosen because they had the lowest Bayesian information criterion, and 30 principal components (PCs) were retained. To corroborate gene pool allocation of genotypes, a simulation of population structure was performed with Structure software (Pritchard et al., 2000) for  $K$ -values of 1 to 10 with the following parameters: burn-in period of 100,000, 100,000 Markov chain Monte Carlo repetitions, and 100 runs. Results were analyzed as described by Evanno et al. (2005), and  $\Delta K$  was plotted with  $K$ -values of 1 to 10, where  $\Delta K$  is the mean of the second-order rate of change of the  $\ln P(D)$  values of a given  $K$  divided by the  $\ln P(D)$  SD (Evanno et al, 2005). Finally, an analysis of molecular variance (AMOVA) was performed using Arlequin 3.5.2.2 software (Excoffier et al., 2005) with marker data of the parents ascribed to the two proposed gene pools to divide the genetic variation into components attributed to the variance within and between proposed pools after excluding SPK004 (from the CIP genebank), a CC that was not part of the parent material at Namulonge.

## RESULTS

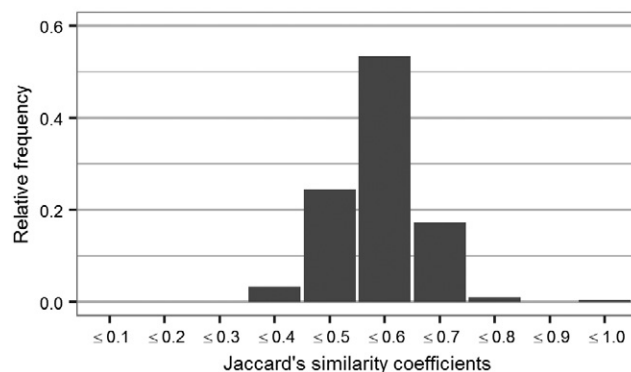
The 31 SSR markers used (Table 2) proved suitable for analyzing the genetic diversity of sweetpotato parents used in the EA breeding platform. A total of 221 alleles were obtained: seven were monomorphic, and 214 were polymorphic (Table 3). On average, 7.13 alleles per locus were obtained. The number of alleles per SSR marker locus ranged from 2 to 12. The average PIC was 0.75. The  $D_L$  was in the range of 0.26 to 0.97, with an average of 0.82. For 14 primer pairs,  $D_L$  was  $>0.90$ .

The Jaccard similarity coefficient and the dendrogram generated with UPGMA revealed large differences among genotypes (Fig. 1 and 2). Jaccard coefficients were in the range of 0.298 to 1.000 (Fig. 1). There was relatively low similarity among most genotypes, with a mean of 0.542 across all pairwise distances. The dendrogram generated (Fig. 2) grouped non-African and African accessions separately (Clusters I and II, respectively). Non-African material clustered according to geographic origin. Four out of six CCs were in Cluster I, whereas CC SPK004 (origin Kenya, from CIP genebank) was in Cluster II and CC Naveto from Papua New Guinea was in an independent subcluster (neither Cluster I nor II), together with clone ‘WT-237’ of unknown origin and two African genotypes. The non-African accessions in Cluster I

**Table 3. Number of alleles, polymorphic information content (PIC), and power of discrimination ( $D_L$ ) for 31 simple sequence repeat loci used to estimate the genetic diversity in parental material from Uganda and checks.**

Primer name	No. of alleles	PIC	$D_L$
Ib-242	5	0.76	0.76
Ib-286	8	0.83	0.95
Ib-297	9	0.82	0.96
IBCIP-1	4	0.75	0.74
IbC5	8	0.79	0.91
IbC12	7	0.84	0.96
IbE2	12	0.8	0.93
IbJ67	8	0.78	0.89
IbJ116a	9	0.81	0.93
IbJ263	5	0.72	0.82
IbJ522a	5	0.75	0.82
IbJ544b	4	0.52	0.26
IbJ664E	4	0.59	0.46
IbN21	9	0.81	0.90
IbN24	4	0.71	0.69
IbN37	12	0.85	0.97
IBS144	9	0.84	0.96
IBS147	9	0.82	0.94
IBS149	12	0.8	0.93
IBS169	5	0.68	0.76
IBS199	11	0.84	0.95
IbY40	7	0.76	0.92
IbY44	8	0.78	0.89
IbY46	8	0.82	0.94
IbY51	6	0.69	0.61
IbY52	6	0.77	0.86
IbY53	5	0.72	0.76
IbY54	6	0.69	0.76
IbY56	2	0.48	0.47
IbY58	5	0.74	0.88
IbY60	9	0.76	0.89
Average	7.13	0.75	0.82

started to merge with African accessions (Cluster II) at a Jaccard coefficient of 0.49. The MVs ‘TIS-9265’ (NG01) and ‘Dagga’ (199062.1) from breeding of IITA and CIP, respectively, were classed in Cluster I but were not part of any subcluster group. The MVs from China (Xushu 18 and Yanshu 1) grouped at a Jaccard coefficient of 0.49, but the MV ‘Tainung 64’ from Taiwan clustered with US varieties. The MVs from the United States—‘Jewel’,



**Fig. 1. Frequency distribution of Jaccard genetic similarity distances among 141 sweetpotato genotypes (135 parents from the East African breeding platform and six check clones).**

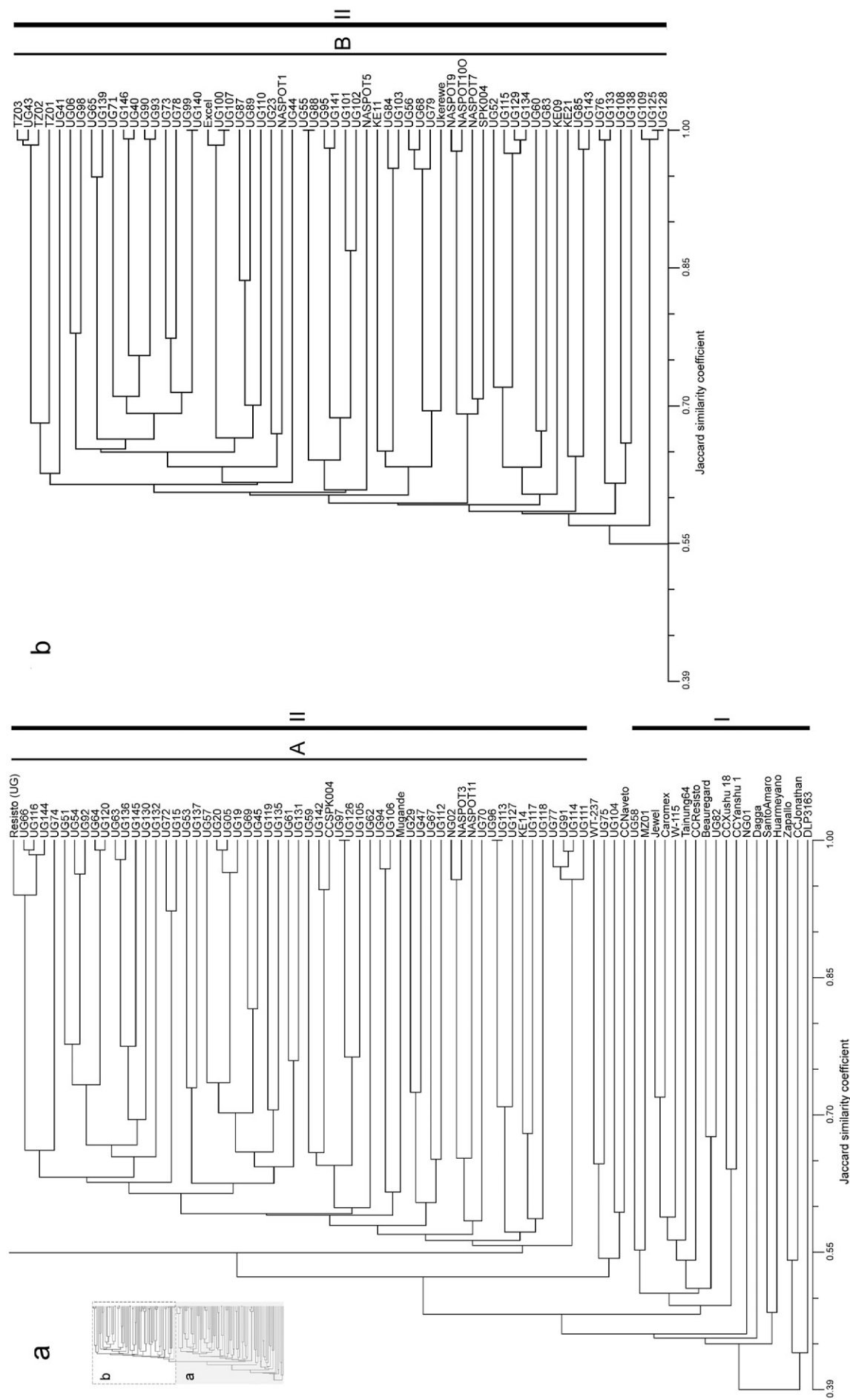


Fig. 2. Dendrogram of the unweighted pair group method analysis based on Jaccard's genetic similarity distances for simple sequence repeats of 135 parents in use at the East African breeding platform and six check clones (labeled by check clone [CC] and variety name).

‘Caromex’, ‘W-115’, CC Resisto (from the CIP genebank), and ‘Beauregard’—grouped consistently. The CC Resisto from the CIP genebank (Cluster I) was very different from the genotype named Resisto (UG) used as a parent in the EA breeding platform (Cluster II).

Most parents from EA (117 clones) grouped in Cluster II (Fig. 2). There was no relationship between grouping of clones and country of origin or district of collection (Table 1, Supplemental Table S1). Six accession pairs in Cluster II were respectively similar (Fig. 2, Supplemental Table S2). At a higher aggregation level, there were two main subclusters in Cluster II at similarity 0.54 (designated Groups A and B in Fig. 2). In Group A of Cluster II (comprising 57 accessions), all genotypes were of EA origin (Fig. 2a). Some smaller subclusters appeared in Group A; for example, at a Jaccard coefficient of 0.95 comprising ‘Dimbuka’ (UG77), ‘Tuulansime’ (UG91), ‘Namusoga’ (UG111), and ‘Bunduguza Empyaka 2’ (UG114), all clones had white- to cream-fleshed root color and pink or purple-red storage root skin color (Supplemental Table S1). Most genotypes in Group A were FVs from different districts of Uganda or different EA countries, and no morphological characteristic distinguished the group. The MVs TIS9101 (NG02) and ‘NASPOT 3’ (UG86) morphologically differed despite their genetic similarity (Supplemental Table S1). The CC SPK004 (from the CIP genebank) and Resisto (UG, a parent in the EA breeding platform) both fell in Group A of Cluster II.

On the basis of SSR marker data, flesh color and storage root shape, the CC SPK004 differed from the SPK004 used as a parent (Group B of Cluster II). Resisto (UG) clustered at high similarity with African clones. Although its orange flesh root color was very close to CC Resisto, on the basis of SSR marker data, plant type, storage root shape, and skin color, Resisto (UG) clearly differed from CC Resisto (Supplemental Table S1). In Group B of Cluster II (comprising 62 accessions), all genotypes were of EA origin, except ‘Excel’ (from the United States). For 11 pairs of genotypes in Group B, Jaccard coefficients were 0.90 to 1.00 (Supplemental Table S2). As for Group A of Cluster II, these pairs in Group B of Cluster II were not geographically associated. All OFSP FVs were found in Group B. Two subclusters in Group B contained OFSP exclusively: the first subcluster comprised Carrot D (TZ01), ‘Mayai’ (TZ02), ‘Carrot C’ (TZ03), ‘Abuket 1’ (UG41), and ‘Ejumula’ (UG43); and the second comprised SPK004 (KE19) and MVs derived from SPK004 (KE19) such as ‘NASPOT 7’ (UG121), ‘NASPOT 9 O’ (UG122), and ‘NASPOT 10 O’ (UG123), all of which are moderately resistant to SPVD.

In further analysis steps, all binary SSR data belonging to genotypes of Cluster I were excluded from the analysis to determine whether Cluster II could be separated into two or more groups (as UPGMA indicated). This was performed in

four analysis steps: (i) PCoA and a biplot of PC score values, (ii) DAPC, (iii) a simulation of population structure, and (iv) AMOVA. The first two PCs for Cluster II genotypes explained 18.1% of the total variation. The biplot of the first two PC axis score values (Fig. 3) separated Groups A and B of Cluster II (Fig. 2). The DAPC separated the accessions into four clusters and explained 80.9% of the variance. The first component separated DAPC Cluster 1 from DAPC Clusters 2 to 4 (Fig. 4). The DAPC Cluster 1 contained the accessions of Group B in Cluster II (Fig. 2b), except for seven clones observed in DAPC Cluster 2. All remaining genotypes in DAPC Clusters 2 to 4 were found in Group A in Cluster II (Fig. 2a), except for ‘Osukut’ (UG51). The DAPC Cluster 3 contained 15 of 17 genotypes found in Group A of Cluster II at the top of the dendrogram (Fig. 2a). The DAPC Cluster 4 contained genotypes from two groups at the bottom of Group A of Cluster II. Simulation of the population structure using Structure (Pritchard et al., 2000) confirmed the PCoA and DAPC results. The  $\Delta K$  obtained for  $K$ -values of 1 to 10 exhibited a clear peak at  $K = 2$  (results not presented). The simulations were plotted for  $K$ -values of 2 to 4 (Fig. 5). For  $K = 2$ , two groups were well differentiated: the first included genotypes of DAPC Cluster 1, and the second included genotypes from DAPC Clusters 2 to 4. For  $K = 3$ , the newly appearing group contained mostly genotypes from DAPC Clusters 3 and 4—this new group in the simulated population structure was represented by two groups within Group A of Cluster II (Fig. 2a). The first group located at a Jaccard coefficient of 0.6 contained 29 accessions, and the second group contained four clones joining at a Jaccard coefficient of 0.94. For  $K = 4$ , similarities between DAPC Clusters 2 and 4 increased. For simulations with  $K > 4$ , the genotypes were allocated in the same pattern, but genotypes from DAPC Cluster 1 started to differentiate into new groups (results not shown).

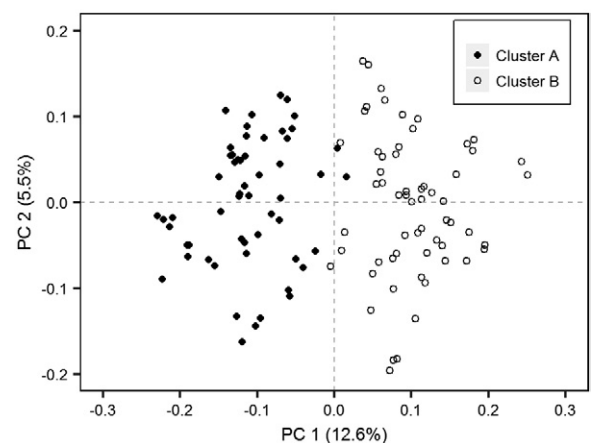


Fig. 3. Principal coordinate analysis based on Jaccard's genetic similarity distances for simple sequence repeats of 118 sweetpotato genotypes from Cluster II and in use as parents at the East African breeding platform (Groups A and B in Fig. 2, except for check clone CC-SPK004, designated as Cluster A and B). PC1 and PC2 are the first and second principal coordinates, respectively.

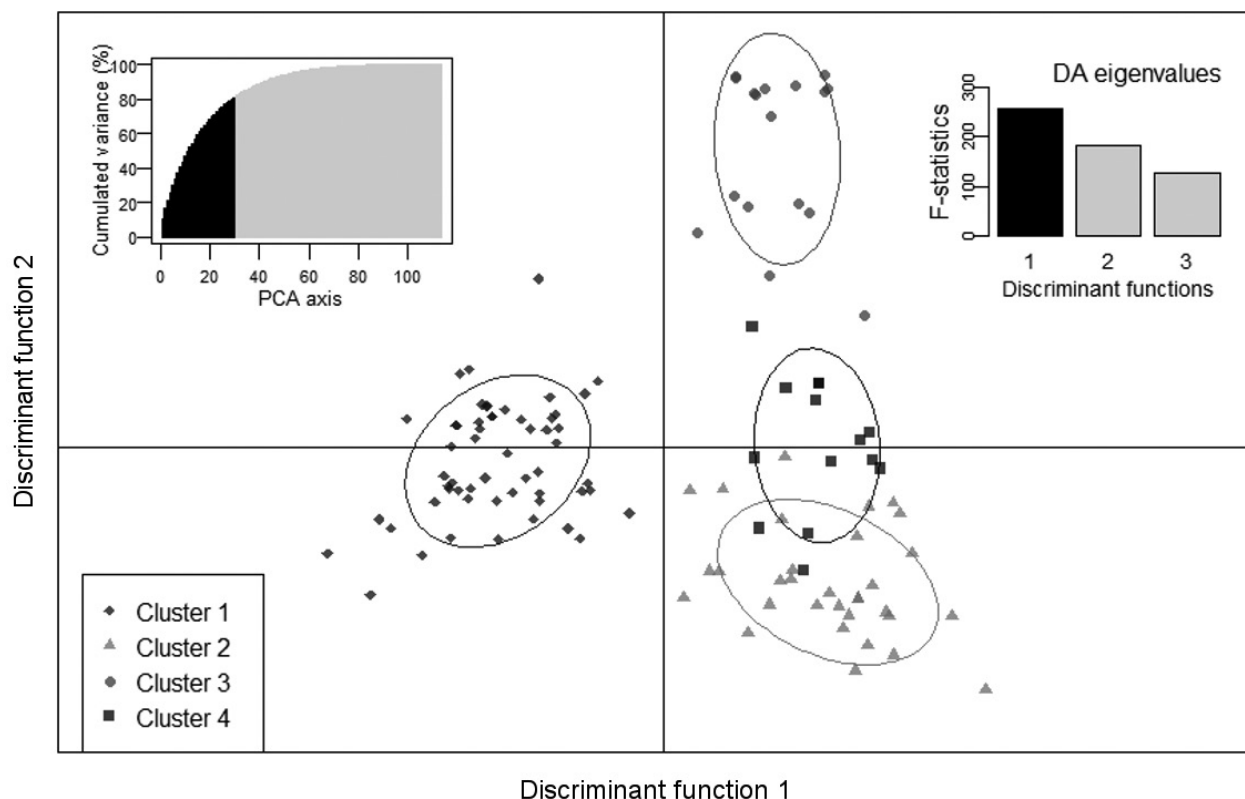


Fig. 4. Discriminant analysis of principal components (DAPC) based on Jaccard's genetic similarity distances for simple sequence repeats of 118 sweetpotato genotypes from Cluster II and in use as parents at the East African breeding platform (Groups A and B in Fig. 2, except for check clone CC-SPK004). The left upper inset graphic shows the cumulative variance by principal component analysis (PCA) axis retained; the right upper inset shows the discriminant analysis (DA) eigenvalues; the left lower inset shows the symbols used for genotypes allocated to DAPC Clusters 1 to 4.

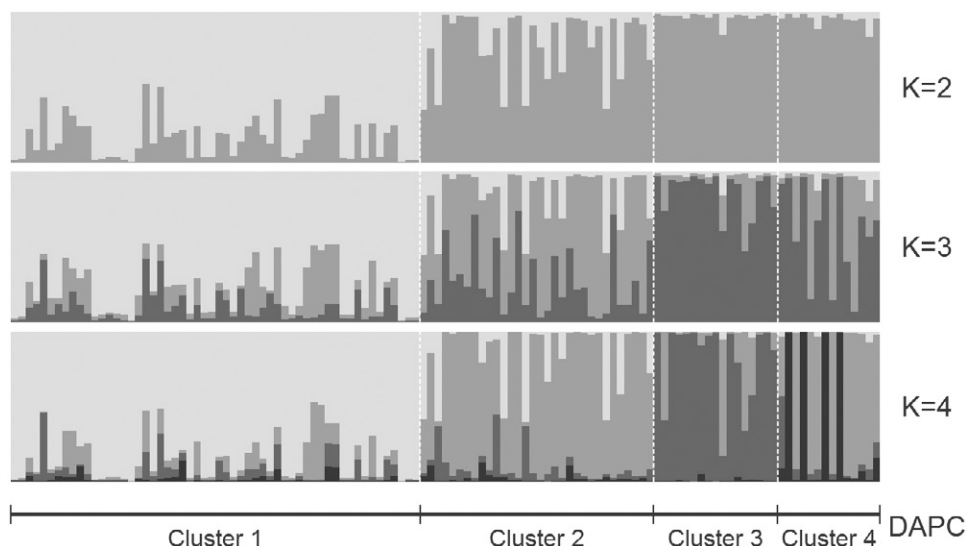


Fig. 5. Bayesian model-based inference of simulated population structure on 118 sweetpotato genotypes from Cluster II and in use as parents at the East African breeding platform (Groups A and B in Fig. 2, except for check clone SPK004). Simulations presented are for K-values of 2 to 4.

With respect to parental material with African origin at Namulonge, two gene pools were proposed (Table 4). The first comprised Group A of Cluster II genotypes (57 clones), and the second comprised Group B of Cluster II genotypes (62 clones) (Fig. 2). Only six African clones

from the crossing block remained unclear concerning gene pool allocation (Fig. 2): 'Bunduguza' (UG75) and 'Uganda Mali' (UG104) in the independent subcluster (neither Cluster I nor II); and 'Rwabuganda' (UG58), 'Kyabafuluki' (UG82), TIS-9265 (NG01), and 'Rainha'



**Table 4. Proposed grouping of parental material in the East African breeding platform.**

Group	No. of genotypes	Clone names (codes)
Group A, Cluster II	57 (56†)	K-566632 (KE14), Mugande (RW01), Magabari (UG05), Karebe (UG15), Kigabali (UG19), Kyebandula (UG20), Osapat (UG29), NN (UG45), Wagabolige (UG47), Osukut (UG51), Kakoba (UG53), Epura Amojong (UG54), Kalobo (UG57), Kibogo (UG59), NN (UG61), Kyebandira 2 (UG62), Anamoyito (UG63), Mary (UG64), Koromojo (UG66), (UG67), Tedolo Kereni (UG69), Burundi (UG70), Kalebe (UG72), NN (UG74), Dimbuka (UG77), NASPOT 3 (UG86), Tuulansime (UG91), Bungoma (UG92), NN (UG94), <u>Koromojo Red (UG96)</u> and <u>Koromojo Red (UG113)†</u> , Bikiramaia (UG97) & Semanda (UG126)‡, NN (UG105), Munafu Dimbuka (UG106), Namusoga (UG111), NN (UG112), Bunduguz Empyaka 2 (UG114), Dimbuka (UG116), New Kawogo (UG117), Bitambi (UG118), Sowola 389A (UG119), NN (UG120), NASPOT 11 (UG124), NN (UG127), NN (UG130), NN (UG131), NN (UG132), Kawogo Old (UG135), NN (UG136), Woluganda (UG137), Tanzania (UG142), Malagalia (UG144), Mbale (UG145), TIS-9101 (NG02), CC SPK004 (CIP441768), and Resisto (UG)
Group B, Cluster II	62	K-118 (KE09), Oguroiwe (KE11), SPK004 (KE19), Ubuogo (KE21), Carrot Dar (TZ01), Mayai (TZ02), Carrot C (TZ03), Ukerewe (TZ04), NN (UG06), Tororo 3 (UG23), Kala (UG40), Abuket 1 (UG41), Ejumula (UG43), Kamamanzi (UG44), Opaade (UG52), <u>Anyumel (UG55)</u> and <u>Dar-es-Salaam Carrot (UG88)‡</u> , Oleke (UG56), NK318L (UG60), Silk Omuyaka (UG65), Liralira (UG68), NN (UG71), Kibanda (UG73), Dimbuka Obuleku (UG76), Silk (UG78), Otada (UG79), NASPOT 1 (UG80), NASPOT 5 (UG81), NASPOT 5/58 (UG83), Kampala Red (UG84), Silk (UG85), NN (UG87), Suwedi (UG89), NN (UG90), Duduma 2 (UG93), Dduka Enzala (UG95), NN (UG98), <u>NN (UG99)</u> and <u>Mpaifumbiro (UG140)‡</u> , NN (UG100) and NN (UG107)‡, Oketodede (UG101), Nylon (UG102), NN (UG103), Mugiga (UG108), Bunduguza 2 (UG109), Kigaire (UG110), Gulu (UG115), NASPOT 7 (UG121), NASPOT 9 O (UG122), NASPOT 10 O (UG123), <u>NK259L (UG125)</u> and <u>Mukoma (UG128)‡</u> , Silimu (UG129), Dimbuka Bukulula (UG133), Dagadaga (UG134), BND145L (UG138), Tengerere (UG139), Mpambire (UG141), Silk (UG143), Tanzania (UG146), and Excel (CIP440016)
Unclear	4	CC Naveto (CIP440131), Bunduguza Empyaka (UG75), Uganda Mali (UG104), and WT-237 (n.a.§)
Exotic material Cluster I	18	Rwabuganda (UG58), Kyabafuluki (UG82), TIS-9265 (NG01), Rainha (MZ01), Santo Amaro (CIP400011), Dagga (CIP199062.1), Huarmeyano (CIP420020), DLP3163 (CIP420269), Zapallo (CIP420027), Tainung 64 (CIP440189), Beauregard (CIP440132), Caromex (CIP440136), Jewel (CIP440031), W-115 (CIP440424), CC Yanshu 1 (440024), CC Xushu 18 (440025), CC Jonathan (CIP420014), and CC Resisto (CIP440001)
Total	141	

† Koromojo Red (UG96 and UG113) with Jaccard's similarity coefficient of 1 (underlined) is truly identical because it was used twice in the crossing block and labeled with different codes, so that the number of clones in Group A is 56.

‡ Pairs of genotypes with Jaccard's similarity coefficient of 1 (underlined), but morphologically different in at least one trait description.

§ n.a., not available.

(MZ01) in Cluster I together with exotic parental material for EA or CCs from North and South America and Asia. For the proposed gene pool allocation, AMOVA revealed 14.0% variation between and 85.0% variation within pools (results not presented; CC SPK004 was excluded from AMOVA because it was not used as a parent at Namulonge). The variation between gene pools was significant ( $P < 0.0001$ ).

For six pairs of genotypes in Cluster II, the Jaccard coefficient was 1.0 (Fig. 2, Table 4, Supplemental Table 2). The pair 'Koromojo Red' (UG96 and UG113) was truly identical because the cultivar was used twice in the crossing block. For the other clone pairs with Jaccard coefficients of 1.0, it remained unclear whether they were truly identical because they were described differently in at least one morphological trait or were not amplified for one primer (Supplemental Tables S1 and S2).

## DISCUSSION

With the objective to test the suitability of HEBS for sweetpotato breeding in SSA, gene pools must be established. It is preferable that these pools be mainly formed by local material. Sweetpotato gene pools have so far only been established at CIP in Peru (Grüneberg et al., 2015). However, increasing genetic gains in SSA has priority at

CIP. There are first steps of using the independent breeding nurseries in southern and northern Mozambique, as well as those in southern and northern Ghana, as separate gene pools, which can be considered as a preliminary gene pool separation. However, for the large polycross seed nursery at Namulonge in Uganda, there is no information that would allow conducting such a preliminary separation of parental material.

Some breeders argue that recombining hexaploid sweetpotato in polycrosses creates a huge amount of genetic variation but that population means for yield and SPVD resistance do not greatly improve. Indeed, there appears to have been almost no improvement in SPVD resistance. Typically, <0.2% of 1000 clones are resistant to SPVD in breeding populations at Namulonge, as was observed almost two decades ago (Mwanga et al., 2002a, 2002b). A HEBS using reciprocal recurrent selection could be attractive for sweetpotato breeders, especially in EA where SPVD pressure is extremely high, because enhanced inbreeding within gene pools is speculated to lead to higher genetic gains for SPVD resistance. Resistance to SPVD is thought to be recessively inherited involving one or two genes (Mwanga et al., 2002a, 2002b) and the frequency of recessive homozygous loci in a hexaploid population is extremely low (Gallais, 2003).

Inbreeding in gene pools, in association with recombination among gene pools to avoid sacrificing heterozygosity, could be a solution.

Our hypothesis that the large number of parents at the breeding platform at Namulonge (135 polycross parents) can be divided into groups was confirmed. As expected, the 10 exotic parents formed a cluster at high aggregation level (Cluster I, Fig. 2; exotic material, Table 4). The parents of African origin formed a separate cluster (Cluster II, Fig. 2) at an aggregation level only slightly lower than the aggregation level comprising exotic parental material and CCs of non-African origin. This African Cluster II was further divided in two groups and gene pools, respectively (Groups A and B, Fig. 1; Groups A and B, Table 4). The aggregation level compared with Cluster I (Fig. 2) indicated a large genetic diversity within Groups A and B of Cluster II, respectively. With respect to genetic diversity, there appears to be little need to introduce non-African clones into the breeding program at Namulonge, and this could even be harmful (e.g., by allowing a gene flow from clones that are highly susceptible to SPVD into EA breeding populations). The cluster pattern observed is consistent with the grouping of CCs (Fig. 2). The five non-African CCs were grouped in Cluster I or in a small independent subcluster, which was neither Cluster I nor II, whereas the CC from Africa (i.e., SPK004 obtained from the CIP genebank) was in Cluster II. SPK004 (CIP441768 held in trust at the CIP genebank) might not be the clone originally thought to be acquisitioned by the genebank. The parent Resisto (UG) at Namulonge appeared to be a local African clone with some striking attributes in common with CC Resisto (CIP440001).

The most striking result of our study was the possibility of proposing for parents of African origin at Namulonge (Cluster II, Fig. 2) two respective groups and gene pools (Table 4), with each pool exhibiting large molecular genetic variation. Distinct groups within adapted germplasm in allogamous crops are raising interest among breeders targeting heterosis and reciprocal recurrent selection schemes. This has been the basis of maize hybrid breeding (Hallauer et al., 2010) and the first step in hybrid breeding for many other crops (Melchinger and Gumber, 1998). To avoid confusion, we propose that in sweetpotato clone breeding the emphasis of gene pool separation and HEBS must be on improved hybrid populations for selection of best clones, and not homogeneous hybrids generated from homozygous inbred lines. However, partially inbred groups or populations in clone breeding could be important in increasing genetic gains for quality and resistance to diseases. A HEBS in sweetpotato breeding comprising A, B, and  $A \times B$  populations might increase the chances of generating more good  $A \times B$  crosses with respect to yield and SPVD resistance. The history of maize hybrid breeding (Hallauer et al., 2010)

shows that maize breeding gradually entered into classical hybrid breeding from observations made on population hybrids (introgression of Northern Flints and Southern Dents), and the population hybrid variety concept in maize is still worth investigating for certain agriculture sectors (Carena, 2005).

An indication of large genetic diversity in our proposed gene pools (Groups A and B, Fig. 2) was that Groups A and B joined at a Jaccard coefficient of  $\sim 0.55$  to form Cluster II. The two groups (Gene Pools A and B) in Cluster II (mainly adapted EA parental breeding material) were more distinct after PCoA (Fig. 3). However, DAPC additionally indicated a potential clustering (Clusters 2–4, Fig. 4) within Group A of Cluster II (Fig. 2a). Simulations by Bayesian model-based inference of population structures (Fig. 5) provided further evidence that grouping Cluster II clones (mainly local parental material) into two breeding pools was the most appropriate gene pool separation (Groups A and B).

The parental material in use (135 clones) at Namulonge comprises mainly local parents (122 clones) and  $\sim 10\%$  exotic parents (13 clones), which should be suitable to develop locally adapted breeding material. Using 141 clones, 135 parental clones, six CCs (Table 1), and 31 SSR primers (Table 2), 214 polymorphic alleles were found with an average of 7.13 alleles per primer (Table 3). Compared with the primary diversity center of sweetpotato (South America, Central America, and the Caribbean), the genetic diversity of the EA breeding platform is indeed large but, as expected, slightly smaller. Roullier et al. (2011) analyzed 329 accessions from South America, Central America, and the Caribbean and reported 4 to 23 alleles per primer using 13 SSR primers (6 of these 13 primers were also used in our study). Some of the SSR primers used in this study (i.e., Ib-242, Ib-286, and Ib-297) were also used by Veasey et al. (2008) in Brazilian germplasm and resulted in similar numbers of polymorphic alleles per primer to those in our study. The same holds true for a germplasm study in Burkina Faso (Koussao et al., 2014) that used two commonly used primers (IBS144 and IBS199). With respect to EA germplasm (seven primers commonly used), higher numbers of alleles were previously reported for some primers (Tumwegamire et al., 2011b): IbJ116a, Ib-242, IbCIP1, IbJ544b, IbJ522a, IbC12, and Ib-297. This might be because Tumwegamire et al. (2011b) used germplasm from more African countries. The set of primers used in our study was highly informative and suitable for analyzing sweetpotato (Table 3). The average PIC (0.75) was larger than the threshold of  $PIC \geq 0.50$  (Botstein et al., 1980). The values of  $D_L$  for the primer set (average  $D_L = 0.82$ , Table 3) represent the power to differentiate a single genotype from an infinite population (Table 3). We consider that IbN37, Ib-297, IbC12,

IBS144, and IBS199 were the most useful primers ( $D_L = 0.95\text{--}0.97$ ).

The Jaccard coefficients with a mean of 0.542 and range of 0.298 to 1.00 were a clear indicator of high genetic variability among the accessions analyzed. The distribution of similarity agrees with previous reports (Zhang et al., 2001; Yada et al., 2010b; Tumwegamire et al., 2011b). It has been argued that large genetic distances between sweetpotato genotypes can be expected even in smaller populations due to hexaploidy, heterozygosity, self-incompatibility, and ease of sexual seed generation (Grüneberg et al., 2015). Other studies also mention propagation by cloning as a driving factor for new genetic diversity in sweetpotato (He et al., 1995; Zhang and Xie, 1998). The AMOVA for the proposed two gene pools revealed lower but significant variation between parents of Groups A and B (14%) than within groups (86%). The amount of molecular variance among parental material of sugar beet breeding programs is much smaller (2.6%; De Riek et al., 2001). Our observed molecular variance between groups and Gene Pools A and B, respectively, was still larger than in studies for seven tropical maize populations (Reif et al., 2003), with 10.2% between-population variation, and soybean [*Glycine max* (L.) Merr.] landraces from three West Pacific countries with 12.4% variation among countries (Li and Nelson, 2001).

With respect to duplicates among the parental material in use at Namulonge, only one pair of accessions was certainly duplicated (Koromojo Red, UG96, and UG113, Group A of Cluster II in Fig. 2 and Table 4), which reduces the number of parents in the crossing block of Namulonge to 134. Some accessions could be duplicates on the basis of the Jaccard coefficient of one (five pairs indicated in Table 4), but accessions in these pairs were described differently for at least one morphological trait (Supplemental Table S1) or did not amplify for one primer. Currently, we suggest that all five clone pairs remain as parents in the crossing block. With this treatment of the “duplicates” identified in our study, Gene Pools A and B comprised 56 and 62 parents, respectively (Table 4). This number of parents should be sufficient to test whether Gene Pools A and B are mutually heterotic and to discard a larger number of parents (40–50%) in each gene pool on the basis of poor offspring performance concerning yield and SPVD susceptibility. We expect that this effort will result in significant improvement of the breeding population at the East African sweetpotato breeding platform in Namulonge. Moreover, this presents the option to repeat elite cross combinations on a large scale for National Agriculture Research System partners. A remaining critical issue is how exotic and unclearly grouped parents (17 clones excluding CCs, Table 4) should be treated in the breeding platform at Namulonge. Genetic diversity

in the proposed Gene Pools A and B is certainly large. The exotic and unclearly grouped parents might not be needed. Only in cases for which breeders consider that the parents have valuable attributes not present as alleles in Gene Pools A and B should the exotic material be kept apart to establish a separate prebreeding program with local EA parents.

Our study showed a successful separation of gene pools in a large set of parental sweetpotato material using SSR markers. Further studies are required to test the hypothesis that the gene pools are mutually heterotic. Our study should help in the search for gene pools in other clonally propagated crops for testing of HEBS.

## Conflict of Interest

The authors declare that there is no conflict of interest.

## Supplemental Material

Supplemental tables for this article are available online. Tables include a description of clones used for the genetic diversity study (parental material and checks; Supplemental Table S1), and specific groupings of parental material based on Jaccard's SSR and UPGMA (Supplemental Table S2).

## Acknowledgments

This study was supported by grants from the Bill & Melinda Gates Foundation and the CGIAR Research Program on Roots, Tubers and Bananas

## References

- Becker, H.C., G.M. Engqvist, and B. Karlsson. 1995. Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers. *Theor. Appl. Genet.* 91:62–67. doi:10.1007/BF00220859
- Bhosale, S.U., B. Stich, H. Frederick, W. Rattunde, E. Weltzien, B.I.G. Haussmann, et al. 2011. Population structure in sorghum accessions from West Africa differing in race and maturity class. *Genetica (The Hague)* 139:453–463.
- Bosemark, N.O. 2006. Genetics and breeding. In: A.P. Draycott, editor, *Sugar beet*. Blackwell Publ., Oxford, UK. p. 50–88. doi:10.1002/9780470751114.ch4
- Botstein, D., R.L. White, M. Skolnick, and R.W. Davies. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am. J. Hum. Genet.* 32:314–331.
- Buteler, M.I., R.L. Jarret, and D.R. Labonte. 1999. Sequence characterization of microsatellites in diploid and polyploid *Ipomoea*. *Theor. Appl. Genet.* 99:123–132. doi:10.1007/s001220051216
- Carena, M.J. 2005. Maize commercial hybrids compared to improved population hybrids for grain yield and agronomic performance. *Euphytica* 141:201–208. doi:10.1007/s10681-005-7072-0
- De Riek, J., E. Calsyn, I. Everaert, and E. Van Bockstaele. 2001. AFLP based alternatives for assessment of distinctness, uniformity and stability of sugar beet varieties. *Theor. Appl. Genet.* 103:1254–1265. doi:10.1007/s001220100710



- Elameen, A., S. Fjellheim, A. Larsen, O.A. Rognli, L. Sundheim, S. Msolla, et al. 2008. Analysis of genetic diversity in a sweet potato (*Ipomoea batatas* L.) germplasm collection from Tanzania as revealed by AFLP. *Genet. Resour. Crop Evol.* 55:397–408. doi:10.1007/s10722-007-9247-0
- Elameen, A., A. Larsen, S.S. Fjellheim, L. Sundheim, S. Msolla, E. Masumba, and O.A. Rognli. 2011. Phenotypic diversity of plant morphological and root descriptor traits within a sweet potato, *Ipomoea batatas* (L.) Lam., germplasm collection from Tanzania. *Genet. Resour. Crop Evol.* 58:397–407. doi:10.1007/s10722-010-9585-1
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x
- Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1:47–50. doi:10.1177/117693430500100003
- FAOSTAT. 2018. FAOSTAT database. FAO, Rome. <http://faostat3.fao.org/faostat-gateway/go/to/download/Q/QC/E> (accessed 22 June 2018).
- Fischer, S., A.E. Melchinger, V. Korzun, P. Wilde, B. Schmiedchen, J. Möhring, et al. 2010. Molecular marker assisted broadening of the Central European heterotic groups in rye with Eastern European germplasm. *Theor. Appl. Genet.* 120:291–299. doi:10.1007/s00122-009-1124-0
- Gallais, A. 2003. Quantitative genetics and selection theory in autopolyploid plants. *Inst. Natl. Rech. Agron., Paris*.
- Gichuru, V., V. Aritua, G.W. Lubega, R. Edema, E. Aadipala, and P.R. Rubaihayo. 2006. A preliminary analysis of diversity among East African sweetpotato land races using morphological and simple sequence repeats (SSR) markers. *Acta Hortic.* 703:159–164. doi:10.17660/ActaHortic.2006.703.18
- Grüneberg, W.J., D. Ma, R.O.M. Mwanga, E.E. Carey, K. Huamani, F. Diaz, et al. 2015. Advances in sweetpotato breeding from 1992 to 2012. In: J. Low et al., editors, *Potato and sweetpotato in Africa: Transforming the value chains for food and nutrition security*. CAB Int. p. 3–68. doi:10.1079/9781780644202.0003
- Grüneberg, W.J., R.O.M. Mwanga, M. Andrade, and H. Dapaah. 2009. Sweetpotato breeding In: *Unleashing the potential of sweetpotato in sub-Saharan Africa: Current challenges and way forward*. CIP, Nairobi, Kenya. p. 1–42.
- Hallauer, A.R., M.J. Carena, and J.B. Mianda Fiho. 2010. Quantitative genetics in maize breeding. *Handb. Plant Breed.* Vol. 6. Springer, New York. doi:10.1007/978-1-4419-0766-0
- He, G., C.S. Prakash, and R.L. Jarret. 1995. Analysis of genetic diversity in sweetpotato (*Ipomoea batatas*) germplasm collection using DNA amplification fingerprinting. *Genome* 38:938–945.
- Hull, F.H. 1945. Recurrent selection for specific combining ability in corn. *J. Am. Soc. Agron.* 37:134–145. doi:10.2134/agronj1945.00021962003700020006x
- Jansky, S.H., A.O. Charkowski, D.S. Douches, G. Gusmini, C. Richael, P.C. Bethke, et al. 2016. Reinventing potato as a diploid inbred line-based crop. *Crop Sci.* 56:1412–1422. doi:10.2135/cropsci2015.12.0740
- Jombart, T. 2008. Adegnet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405. doi:10.1093/bioinformatics/btn129
- Kaledzi, P.D., E. Ansong, H.K. Dapaah, and G. Timpo. 2010. Agro-morphological characterization of sweet potato (*Ipomoea batatas*) germplasm. *Ghana J. Hortic.* 8:1–11.
- Kamvar, Z.N., J.F. Tabima, and N.J. Grünwald. 2014. Poppr: A R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281. doi:10.7717/peerj.281
- Kloosterman, A.D., B. Budowle, and P. Daselaar. 1993. PCR-amplification and detection of the human D1S80 VNTR locus. *Int. J. Legal Med.* 105:257–264. doi:10.1007/BF01370382
- Koussao, S., V. Gracen, I. Asante, E.Y. Danquah, J.T. Ouedraogo, T.J. Baptiste, et al. 2014. Diversity analysis of sweet potato (*Ipomoea batatas* [L.] Lam) germplasm from Burkina Faso using morphological and simple sequence repeats markers. *Afr. J. Biotechnol.* 13:729–742. doi:10.5897/AJB2013.13234
- Li, Z., and R.L. Nelson. 2001. Genetic diversity among soybean accessions from three countries measured by RAPDs. *Crop Sci.* 41:1337–1347. doi:10.2135/cropsci2001.4141337x
- LI-COR. 2005. Saga<sup>GT</sup> automated microsatellite analysis software user guide. LI-COR, Lincoln, NE.
- Lindhout, P., D. Meijer, T. Schotte, R.C.B. Hutten, R.G.F. Visser, and H.J. van Eck. 2011. Towards F<sub>1</sub> hybrid seed potato breeding. *Potato Res.* 54:301–312. doi:10.1007/s11540-011-9196-z
- Link, W., C. Dixkens, M. Singh, M. Schwall, and A.E. Melchinger. 1995. Genetic diversity in European and Mediterranean faba bean germplasm revealed by RAPD markers. *Theor. Appl. Genet.* 90:27–32. doi:10.1007/BF00220992
- Low, J., J. Lynam, B. Lemaga, C. Crissman, I. Barker, G. Thiele, et al. 2009. Sweetpotato in sub-Saharan Africa. In: G. Loebenstein and G. Tottappilly, editors, *The sweetpotato*. Springer, New York. p. 359–390.
- Low, J., T. Walker, and R. Hijmans. 2001. The potential impact of orange-fleshed sweet potato on Vitamin-A intake in sub-Saharan Africa. Paper presented at the Regional Workshop on Food-Based Approaches to Human Nutritional Deficiencies: The VITAA Project, vitamin A and orange-fleshed sweet potato in sub-Saharan Africa, Nairobi, Kenya. 9–11 May 2001. *Int. Potato Ctr.*, Lima, Peru.
- Martin, F.M., and A. Jones. 1986. Breeding sweetpotatoes. *Plant Breed. Rev.* 4:313–345.
- Melchinger, A.E. 1999. Genetic diversity and heterosis. In: J.G. Coors and S. Pandey, editors, *The genetics and exploitation of heterosis in crops*. ASA, Madison, WI. p. 99–118. doi:10.2134/1999.geneticsandexploitation.c10
- Melchinger, A.E., and R.K. Gumber. 1998. Overview of heterosis and heterotic groups in agronomic crops. In: K.R. Larnkey and J.E. Staub, editors, *Concepts and breeding of heterosis in crop plants*. Spec. Publ. 25. CSSA, Madison, WI. p. 29–44. doi:10.2135/cssaspecpub25.c3
- Messmer, M.M., A.E. Melchinger, J. Boppenmaier, E. Brunklaus-Jung, and R.G. Herrmann. 1992. Relationship among early European maize inbreds: Genetic diversity among flint and dent lines revealed by RFLPs. *Crop Sci.* 32:1301–1309. doi:10.2135/cropsci1992.0011183X003200060001x
- Miles, J.W. 2007. Apomixis for cultivar development in tropical forage grasses. *Crop Sci.* 47:S238–S249. doi:10.2135/cropsci2007.04.0016IPBS
- Murray, M.G., and W.F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8:4321–4326. doi:10.1093/nar/8.19.4321
- Mwanga, R.O.M., M.I. Andrade, E.E. Carey, J.W. Low, G.C. Yencho, and W.J. Grüneberg. 2017. Sweetpotato (*Ipomoea batatas* L.). In: H. Campos and P.D.S. Caligari, editors, *Genetic improvement of tropical crops*. Springer Int., Cham, Switzerland. p. 181–218. doi:10.1007/978-3-319-59819-2\_6



- Mwanga, R.O.M., J. Moyer, D. Zhang, E.E. Carey, and G.C. Yencho. 2002a. Nature of resistance to sweetpotato virus diseases. *Acta Hort.* 583:113–119. doi:10.17660/ActaHortic.2002.583.12
- Mwanga, R.O.M., C.G.C. Yencho, and J.W. Moyer. 2002b. Diallel analysis of sweetpotatoes for resistance to sweetpotato virus disease. *Euphytica* 128:237–248. doi:10.1023/A:1020828421757
- Perrier, X., and J.P. Jacquemoud-Collet. 2006. DARwin software. Ctr. Coop. Int. Rech. Agron. D  v., Paris. <http://darwin.cirad.fr/darwin> (accessed 16 Oct. 2014).
- Pfeiffer, W.H., and B. McClafferty. 2007. HarvestPlus: Breeding crops for better nutrition. *Crop Sci.* 47:S-88–S-105. doi:10.2135/cropsci2007.09.0020IPBS
- Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- R Core Team. 2014. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna, Austria.
- Reif, J.C., A.R. Hallauer, and A.E. Melchinger. 2005. Heterosis and heterotic patterns in maize. *Maydica* 50:215–223.
- Reif, J.C., A.E. Melchinger, X.C. Xia, M.L. Warburton, D.A. Hoisington, S.K. Vasal, et al. 2003. Genetic distance based on simple sequence repeats and heterosis in tropical maize population. *Crop Sci.* 43:1275–1282. doi:10.2135/cropsci2003.1275
- Rohlf, F. 2000. NTSYS-pc, Numerical taxonomy and multivariate analysis system. Version 2.1. Exeter Publ., New York.
- Roullier, C., G. Rossel, D. Tay, D. McKey, and V. Lebot. 2011. Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. *Mol. Ecol.* 20:3963–3977. doi:10.1111/j.1365-294X.2011.05229.x
- Schaffeitner, R., L.R. Tincopa, O. Palomino, G. Rossel, R.F. Robles, R. Alagon, et al. 2010. A sweetpotato gene index established by de novo assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers. *BMC Genomics* 11:604. doi:10.1186/1471-2164-11-604
- Tams, S.H., E. Bauer, G. Oettler, and A.E. Melchinger. 2004. Genetic diversity in European winter triticale determined with SSR markers and coancestry coefficient. *Theor. Appl. Genet.* 108:1385–1391. doi:10.1007/s00122-003-1552-1
- Tumwegamire, S., R. Kapinga, P.R. Rubaihayo, D.R. Labonte, W.J. Gr  neberg, G. Burgos, et al. 2011a. Evaluation of dry matter, protein, starch, sucrose,  $\beta$ -carotene, iron, zinc, calcium, and magnesium in East African sweetpotato [*Ipomoea batatas* (L.) Lam] germplasm. *HortScience* 46:348–357.
- Tumwegamire, S., P.R. Rubaihayo, D.R. LaBonte, F. Diaz, R. Kapinga, E.O.M. Mwanga, and W.J. Gr  neberg. 2011b. Genetic diversity in white- and orange-fleshed sweetpotato farmer varieties from East Africa evaluated by simple sequence repeat markers. *Crop Sci.* 51:1132–1142. doi:10.2135/cropsci2010.07.0407
- Veasey, E., A. Borges, M.S. Rosa, and J.R. Queiroz-Silva. 2008. Genetic diversity in Brazilian sweet potato (*Ipomoea batatas* (L.) Lam., Solanales, Convolvulaceae) landraces assessed with microsatellite markers. *Genet. Mol. Biol.* 31:725–733. doi:10.1590/S1415-47572008000400020
- Weir, B.S. 1996. Genetic data analysis II. 2nd ed. Sinauer Assoc., Sunderland, MA.
- Xiao, J., J. Li, L. Yuan, S.R. McCouch, and S.D. Tanksley. 1996. Genetic diversity and its relationship to hybrid performance and heterosis in rice revealed by PCR-based markers. *Theor. Appl. Genet.* 92:637–643. doi:10.1007/BF00226083
- Yada, B., P. Tukamuhabwa, A. Alajo, and R.O.M. Mwanga. 2010a. Morphological characterization of Ugandan sweetpotato germplasm. *Crop Sci.* 50:2364–2371. doi:10.2135/cropsci2009.04.0199
- Yada, B., P. Tukamuhabwa, B. Wanjala, D.J. Kim, R.A. Skilton, A. Alajo, and R.O.M. Mwanga. 2010b. Characterization of Ugandan sweetpotato germplasm using fluorescent labeled simple sequence repeat markers. *HortScience* 45:225–230.
- Y  nez, A.V.O. 2002. Aislamiento y caracterizaci  n de marcadores moleculares microsat  lites a partir de la construcci  n de librer  as gen  micas enriquecidas de camote (*Ipomoea batatas* (L) Lam). B.S. thesis, Univ. Nac. Mayor San Marcos, Lima, Peru.
- Zhang, D.P., D. Carbajulca, L. Ojeda, G. Rossel, S. Milla, C. Herrera, and M. Ghislain. 2001. Microsatellite analysis of genetic diversity in sweetpotato varieties from Latin America. In: CIP Program Report 1999–2000. Int. Potato Ctr., Lima, Peru. p. 295–301.
- Zhang, L.Y., and Y.Z. Xie. 1998. Inheritance of flesh colour and its correlation with other traits in sweet potato (*Ipomoea batatas*). *J. Agric. Sci.* 4:30–34.